STORET/WQX Conference Call – August 22nd, 2013 12:00- 1:00 P.M. Eastern Time

Introduction (Charles Kovatch)

- The minutes from all previous conference calls are available over the web:
  http://www.epa.gov/storet/confcalls.html
- The next scheduled call will be September 2013. The exact date will be emailed via the list
  server when the call gets closer.
- Please email STORET@epa.gov and let EPA know you attended the call so that meeting rosters
  may be kept.
- If you have a special topic you would like to lead for an upcoming call, please email
  Kovatch.Charles@epamail.epa.gov.
- EPA would like to hear comments you have on the quality of these conference calls. Please send
  them to STORET@epa.gov
- Please subscribe to the STORET automated server for announcements regarding conference
  calls: http://www.epa.gov/storet/listserv.html

Data Quality Levels (Michael Brennan)

   Today's discussion focuses on water quality monitoring data quality levels. The level of quality
associated with a result is subjective, and is driven by the extent of metadata and quality assurance
associated with a specific result. Metadata, information found in a project QUAPP, is data about data,
such as detection levels, analytical method, precision, bias, and dilution. The organization structure of
WQX includes fields that document a records metadata and level of quality assurance such as fields for
method detection level and precision.

Currently data submitters can classify data as meeting a certain data quality level in several ways.

- Documenting the data quality level in the result comment field
- Documenting the data quality level in the activity comment field
- Documenting levels of data quality at the Project level.   One activity can be associated to
  multiple projects i.e. the same activity ID can be linked to a routine monitoring data project, a
  Toxicity data project, and a data quality level 3 project.
-

Open discussion on classifying data according to a certain level of data quality.

Chris Nuemiller (Washington State)

- Washington has an AIM system indicate data quality at the study level, everything within the
  study is ranked. Made modifications, needs for data to be ranked according to individual pieces
  of data within the study   currently use the EPA 1, 2a 3 rankings within a system, gives them the
  flexibility to add more. for individual Washington's data quality flags results at the study level.

Everything within a study is ranked based on planning and a final result – modifications within a system. Modifications within a study. Data ranked individual pieces within a study added a field that captured that. Currently use EPA 1, 2a , 3 rankings within that field provides flexibility to add more.

Dwane Young (USEPA)

- One of the challenges we have come across at EPA , is determining where to put the QA flag. Are you putting the data quality flag at the sample level or the individual result level.

Chris Nuemiller (Washington State)

- At the result level. At the project level a flag is placed to direct a user to the result level

Dwane Young (EPA)

- What drives that ranking at the result level? Is it related to the method, holding time?

Chris Nuemiller (Washington State)

- At the moment, Third party validation. In cases where you have a parameter such as a poly aromatic hydrocarbon (PAH) within a sample that is looked at with more scrutiny.   The PAH may have a different data quality ranking than the rest of the results in a sample.   At the moment sediment data is using the EPA ranking exclusively, with the flexability to add additional fields for validation and QA.

Link to Washington QA URL: http://www.epa.gov/storet/download/example.xlsx

Sue McCarthy (EPA Region 10)

- We discussed previously adding a field for this purpose. The EPA Superfund program had designated qualifiers of the qualifiers, clarifying whether data went through automated QA, hands on QA or No QA. Similar to what Washington has. Determining whether the value has been validated and what level of validation does it have.

Link to EPA Regoin 10 QA

http://www.epa.gov/storet/download/APPENDIXBDataElementDictionary08062013.xlsx

Chris Nuemiller (Washington State)

- What region 10 is referring to is assigned at the site level with their data. Sometimes datasets are not ranked at the same level so Washington has developed the ability to assign data quality at the result level. Up to the user to rank it who's submitting the data.

Revitale Katznelson (California)

- Important to separate actual error associated with the data from the level of validation. Basic result of data validation process – did the system work, is it ok or is it broken. Qualifier codes such as j's, or r's are flags for measurement systems that are not working properly. Whatever is wrong with the data, it is not detrimental that we use it, but need to use data with caution.
- It's also important to have a sense of the accuracy and precision of values. What kind of accuracy and precision is appropriate, did it meet data quality objectives or not. I don't know if you are interested in the accuracy of data being 1% or 5% may be useful for data miner to know how much error is associated with a specific result.
- It's important to put rankings or color codes on the data to determine has it been QA'd or not. If it has not been QA'd we need to know that
- Data elements process of data temporary vs final can be expanded to determine what level of validation, was taken.
- Whole slough of chemistry data qualifiers - validation status – status of data.

Dwane Young (USEPA)

- There are currently a whole suite of data elements to document accuracy and precision, . Not sure how much they are used but they are there.
- With this discussion we want to address instances where users would like to use want to use outside party data for 303d quality assessment, court proceedings etc. Certain agencies have a set of data rigor rules, data quality objectives established that must be met in order for the agency to use third party data. For example, an agency may outline a criteria of 10 elements that must be met in order for data to be considered top level.   Once data is classified as top level it can be used for legal purposes. Alternatively if only 8 of the elements are met of the outlined criteria, the data may only be used for screening purposes.
- We would like to hear from the user community how this level of screening is being conducted.
- How do we exchange this level of information? Is a data quality level data component something we need in the schema ?
- How are you doing interstate data assessments, at the project level or the result level.

Chris Nuemiller (Washington State)

- Washington ranks their data on a scale of 1-5.
- A data quality level of 1 indicating- Data not verified or assessed for usability.
- A data quality level of 5.-data verified and assessed for usability.   Peer reviewed study report.
- Washington doesn't   use data below 2.

- Who makes the distinction? Data submitter. Data submitter decides when they submit the data. Result validation level for each protocol assured of the data.

Tim Bowren   (Indiana)

- At this point, Indiana is only using data collected by Indiana Department of Environmental Management for state assessments.   IDEM, will be accepting external three data levels 1,2,3 in the future.   IDEM accepts third party data and enters data into their database, subsequently making the determination on data quality themselves.

Dwane Young (EPA)

- Is there a way to provide an additional data quality component in an easy and succinct way that would make sense to folks

Tracie Merril (Seldovia Village Tribe)

- Data quality based on EPA approved QUAPP. For environmental sampling and monitoring. Won't put data in that's bad data.   Will provide additional data in comment field and not provide any data rankings.

Arne Sjordin (Colorado)

- Colorado Working on processes for third party data. Implementing a defined methodology for flagging data. Colorado doesn't host third party data, a lot of it goes to data sharing network.

Dwane Young (EPA)

- There is no short hand way of saying this data is of a certain quality. To address this currently, we have all the data elements in the schema that could be used to describe data quality, collection methods, analytical methods information in the QUAPP.
- We have considered describing the quality of a data record at the project level. For example, a dataset sampled from a specific river could have one sample of questionable quality. This data set could be split into two or more projects to differentiate between values of high quality and more questionable quality.
- We could add a data element at the project level or at a more atomic level like the result level or activity level.

Revitale Katznelson (California)

- Every data piece has its own level of Mishigas associated with it. Results can have different implications depending on the analyst, equipment or analytical method.
- Third party mining the database needs a quick and easy way to tell whats wrong with the data, I don't know if you can get in one data field to satisfy everyone, should be at result level.

- Data field list of allowable values or user defined becomes variable if its user defined

Arne (Colorado)

- set of user defined qualifiers would be preferable, that would put the onus on the data submitter, to fit it within the framework .

Revitale Katznelson (California)

- Agree that is should be a list of allowable values, list of user defined values is almost like a comment field. You would need a dictionary to define them. May be beneficial to have field classified as "other" to alert a data miner that there is an issue with the data that does not fit within any of the defined categories.

Sue McCarthy (USEPA region 10)

- In regards to lab deliverables help to define the data, agree that this should be at the result level.
- Using the Water quality portal, users are not able to retrieve result level information.
- Currently use comment field for qualifier information

Dwane Young (EPA Headquarters)

- this is the beginning of the discussion, the intent is to gather some information to begin this discussion.
- We want to avoid changing the schema
- We may want to build this change in so its backward compatible
- There still is the question of what to do with existing data in the warehouse.

Sue McCarthy (USEPA region 10)

- Repurposing field to say not validated prior data we don't know about.

Tim Bowren   (Indiana)

- Prefer a set of defined qualifiers. clearest interpretation for anyone that decides to mine data at a later point.

Bob Simpson (EPA region 3)

- Earthsoft has been working to build a bridge between SEDD and WQX. May want to bring them into the conversation in regards to the data elements that currently exist.

Participants on the call (based on who emailed STORET@epa.gov)

Daniela Gurlin	Nebraska Department of Environmental Protection
Tracie Merrill	Seldovia Village Tribe
Melanie Titus	New Hampshire department of Environmental Protection
Donald Kean	North Carolina Department of Natural Resources
Tell Judkins	Iowa Tribe of Oklahoma
Debbie Dotson	Eastern Shawnee Tribe of Oklahoma
Arlene Garcia	Pueblo of Acoma
Samantha Bennett	EarthSoft
Shannon Minerich	South Dakota Department of Natural Resources
Jonathan Burian	EPA Region 5
Susanne Meidel	ME Department of Environmental Protection
Michael J. Whitman	West Virginia Department of Environmental Protection
Melissa Miracle	Kentucky Division of Water
Joseph L. Gross	North Dakota Division Of Water Quality
Chris Nuemiller	Washington Department of Ecology
Diane Stevenson	Cherokee Nation Environmental Programs
Andrea Thomas	North Carolina Department of Water Quality
Aaron Leingang	South Dakota Natural Resources
Tim Bowren	Indiana Department of Environmental Management
Arne Sjodin	Colorado Dept. of Public Health & Environment
James F. Hudson	Wisconsin Department of Natural Resources

Lowell Carty	**Arizona Department of Environmental Quality**
Ben Cole	**Maryland Department of Natural Resources**
Jim Porter	**Minnesota Department of Natural Resources**
Caitlyn Nichols	**NEIWPCC**
Michael Canova	**USGS**
Lemonteh' Horne	**Florida Department of Environmental Protection**
Lara Panayotoff	**Kentucky Department for Environmental Protection**
Scottie Wallace	**Mississippi Band Of Choctaw Indians**
Sue McCarthy	**USEPA Region 10**